MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231008

Introducing a Comprehensive Score of Systemic Anticancer Treatment Relevance

Travis ZACK[a] and Jeremy L. WARNER[b,1]
[a] *University of California, San Francisco, CA, USA*
[b] *Brown University and Lifespan Cancer Institute, Providence, RI, USA*
ORCiD ID: Travis Zack https://orcid.org/0000-0002-1620-6455, Jeremy Warner https://orcid.org/0000-0002-2851-7242

Abstract. Treatment patterns in systemic anticancer therapy (SACT) are extremely varied and complex. While professional society guidelines exist that suggest recommended treatment strategies, these guidelines are produced through an extremely laborious and sometimes opaque manual process, making it impossible for such guidelines to cover all relevant treatment scenarios. To complement these manually curated guidelines, we leveraged a database of 5818 clinical trials and 7012 supporting references from 1943–present to calculate a quantifiable "relevance score". In a pilot evaluation, this score was strongly associated with professional society guideline recommendations, while also providing relevance information on thousands of additional therapies. We show that this score also accurately illustrates trends in SACT adoption over time. We foresee that this score, which comprehensively evaluates the relevance of SACT overall and by cancer subtype, will have utility for clinical practitioners as well as researchers in real-world data.

Keywords. Neoplasms, knowledge representation, relevance

1. Introduction

Cancer treatment is a complex and constantly evolving field of clinical practice, with the pace of drug approvals and clinical trials far outstripping most other fields of medicine [1,2]. This has been accelerated by the emergence of immunotherapy and targeted therapy, which have augmented, and sometimes supplanted, more traditional cytotoxic chemotherapeutics. However, diffusion of new interventions into the clinic is not uniform, with many treatments and protocols first developed decades ago still in widespread use.

Given the rapid and constantly evolving nature of the field since its inception in the early 1940s, we propose an automatically calculated "relevance" score for cancer treatment regimens that could provide high utility. This score is derived from a large and comprehensive cancer clinical trial knowledgebase and takes into account regimen efficacy, prevalence within the field, degree of impact, and recency of utilization within randomized clinical trials (RCTs). Such a score could provide context and evidence for

[1] Corresponding Author: Jeremy L. Warner, MD, MS, FAMIA, FASCO; 593 Eddy St., Providence, RI, USA 02906. +1 (401) 444-4000; jeremy_warner@brown.edu. This work was supported in part by NCI grant U24 CA265879; the funder had no direct role in the work or decision to publish.

practicing oncologists, trainees, clinical cancer informatics researchers, and real-world evidence generators. In the sections that follow, we describe the formulation of the score, initial results, and comparison to human expert-driven guidelines from the National Comprehensive Cancer Network (NCCN), a non-profit organization that issues widely utilized guidelines in the United States and internationally.

2. Methods

Our data source was HemOnc.org, an oncologist-curated database of current and historical standard-of-care treatments in the fields of hematology and oncology with an international scope [3]. This site contains information not just about cancer treatment protocols, but also their levels of evidence, such as dates and results of RCT publications involving each of over 3700 cancer treatment regimens from the 1940s to the present day. Information was extracted from the derivative HemOnc ontology on November 6, 2022; this ontology is available at https://dataverse.harvard.edu/dataverse/HemOnc.

The relevance score was conceptualized across six domains: 1) **efficacy** – was a regimen demonstrably superior in an RCT, with additional points for RCTs with a primary endpoint of overall survival rather than a surrogate endpoint (such as response rate); 2) **toxicity** – was a regimen demonstrably less toxic than a comparator in an RCT; 3) **relative prevalence** – how often was a regimen prospectively studied (as an experimental or control arm, or in the nonrandomized setting) as a fraction of all regimens studied within a defined timeframe, e.g., in the preceding decade; 4) **authority** – has a regimen been published in high-impact journals by authoritative persons, institutions, and/or study groups; has it been studied over a long period of time; was it used as the basis of a regulatory decision, e.g., U.S. Food & Drug Administration (FDA) approval; 5) **availability** – are all the drugs constituting the regimen available for use at the local and/or national levels; and 6) **recency** – how long has it been since results for a given regimen were published in the scientific literature. For the purposes of the current analysis, we only include efficacy, relative prevalence, recency, and authority as measured by publication and FDA approval, due to current underlying limitations of the dataset. We extracted information from each of these domains from the HemOnc ontology and calculated the relevance of each protocol identified in studies conducted since 1943. We calculated global "pan-cancer" relevance and disease-specific relevance, including biomarker-defined subtypes such as EGFR-mutated non-small cell lung cancer (NSCLC).

For comparison of this score to current guidelines, we manually reviewed guidelines from the NCCN for three diseases: 1) Ovarian Cancer Version 5.2022, focusing on high-grade serous ovarian adenocarcinoma; 2) Non-Small Cell Lung Cancer Version 5.2022; and 3) colorectal cancer (CRC), combining Colon Cancer Version 2.2022 and Rectal Cancer Version 3.2022. For each of these diseases, we determined each regimen present within the guidelines. We further stratified each regimen into the highest level of recommendation assigned to each within these guidelines: 0: not mentioned; 1: useful in certain situations/other recommended; 2: preferred. We calculated the mean and standard error for each group for each NCCN disease guideline. To determine the ability for this score to predict inclusion of a regimen within NCCN guidelines, we collapsed all regimens mentioned as either

preferred or useful in certain situations/other and used the R package *pROC* to calculate receiver operator characteristic curve and AUC.

3. Results

After removal of 25 non-cancer conditions (e.g., acquired coagulopathy; sickle cell anemia), there were 5818 studies across 174 distinct cancer conditions. The most studied condition was breast cancer, with 1392 studies, followed by NSCLC, with 673 studies. Across these studies, there were a total of 2935 regimens with sufficient information to calculate a relevance score. When calculated across all cancers, the relevance scores (RS) approximated a log-normal distribution (Figure 1L); the median RS was 6 (10^{th} to 90^{th} percentile: -13 to 58). The most relevant regimens were the immune checkpoint inhibitors pembrolizumab (RS=1542) and nivolumab (RS=1212) monotherapy, and the platinum doublet carboplatin & paclitaxel (RS=594). The former have revolutionized treatment approaches for many solid tumors [4], and the latter is a chemotherapy doublet frequently used across multiple cancer subtypes. The least relevant was aminopterin monotherapy (RS=-61), the first drug shown to induce remission in leukemia but quickly replaced by amethopterin (methotrexate) and never FDA approved [5].

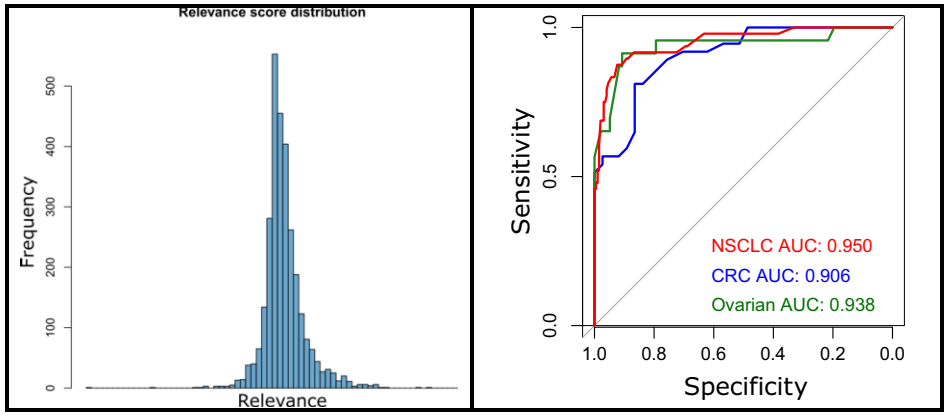

Figure 1. On the left, histogram of n=2935 log-transformed relevance scores, after applying a shift so that all scores are positive; several extreme outliers are not depicted. On the right, ROC curve predicting inclusion of a regimen using RS score for three NCCN guidelines: CRC, NSCLC, and Ovarian.

To compare these results with NCCN disease-specific guidelines, we identified all regimens that were either listed as *Preferred or Other/Recommended* for ovarian NSCLC, and CRC (Table 1). We found very high predictive power of our RS, without inclusion of any additional information (Figure 1R). Additionally, we looked at biomarker-defined subsets of NSCLC, which had 100% concordance with NCCN.

Table 1. Mean regimen relevance score, stratified by NCCN recommendation and disease. Each regimen is only counted in the highest recommendation level found within the NCCN guidelines (some regimens are both preferred and Other/Useful depending on context). Mean, standard errors, and number of regimens are shown.

	Ovarian	**NSCLC**	**CRC**
NCCN Preferred	115+/-22 (n=15)	84 +/-12 (n=34)	69+5.5 (n=30)

NCCN Other/Useful	30+/- 7.8 (n=8)	61+/- 16 (n=14)	60+5.5 (n=7)
Not mentioned by NCCN	0.5+/- 1.4 (n=97)	1.6+/- 1.0 (n=196)	26+/- 3.3 (n=37)
Total	**120**	**244**	**74**

To illustrate how this relevance score can elucidate the evolution of disease treatment as a field grows and changes, we applied our relevance scoring retroactively in two different use scenarios: CRC from 1987-2022 (Figure 2L) and diffuse large B-cell lymphoma (DLBCL) from 1980-2022 (Figure 2R).

In DLBCL we see the dominance of CHOP disrupted by R-CHOP in the early 2000s, based on the seminal LNH 98-5 trial [6], followed by the very recent emergence of chimeric antigen receptor T-cell therapy (CAR-T). In CRC, we see that the relevance of FULV and irinotecan monotherapy, single-agent chemotherapy options, has gradually been replaced by several different doublet and triplet options that are considered to be the standard of care in 2022.

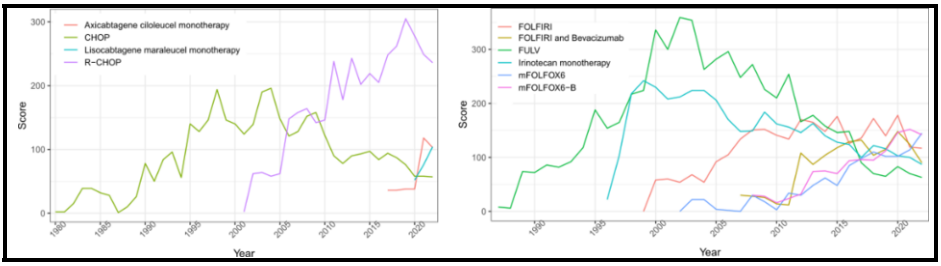

Figure 2. Temporal changes in relevance scores for selected regimens in DLBCL (left) and CRC (right). In DLBCL, CHOP is shown to be supplanted by R-CHOP with recent increasing relevance of two CAR-T therapies. In CRC, monotherapies FULV and irinotecan have decreased in relevance as chemotherapy doublets/triplets, e.g., FOLFIRI and mFOLFOX6 have increased.

4. Discussion

The set of approved treatment regimens in clinical oncology is exponentially expanding. New approaches don't completely supplant previous ones, but instead add to a growing catalog of treatment options. Herein we introduce a context-specific way of assessing a treatment's current and past relevance to a specific disease or context, powered by a large, rich dataset of oncological clinical trials and publications. This tool uses the recency, frequency, efficacy, and impact of each regimen in clinical trials and resulting publications to quantify the relevance of each regimen in the zeitgeist of a disease. We demonstrate this score to be highly predictive of guideline inclusion and demonstrate several ways that such a score could be useful.

Oncology societies release clinical guidelines with often adopted as "standard of care". Due to the fields complexity, frequent treatment intolerance, and increasing sequential lines of therapy, these guidelines can't cover every scenario. Guidelines also take a great deal of time and effort to produce and may become outdated quickly. A comprehensive metric that captures the relevance of all potential treatments, and can be updated automatically, can be a useful reference for providers, medical staff, and clinical researchers. The ability to report this score as a function of time also allows perspective within retrospective real-world data analysis of oncology practices.

There are some notable limitations in this work. Often very similar regimens are used within a disease, such as FOLFOX4 vs mFOLFOX6. Ideally, one could weight and combine evidence in cases of very similar regimens to avoid bifurcation that may lead to lower individual scores; relevance network approaches may solve this issue [7]. There is also some missingness in the utilized dataset, primarily data regarding years of enrollment for non-randomized studies, toxicity, and drug availability outside of the FDA domain. We also found that impact factor of the journal where the study is published provided significant benefit to incorporation within our score but recognize that impact factor can vary widely over time, which we are not incorporating into our model currently. Finally, we recognize that this score is not a substitute for clinical guidelines compiled by experts within the subfields of oncology, nor is that its intention. Rather we see this as a flexible, transparent, and highly accessible supplement to understand the history and current state of disease treatment in the many subfields of oncology.

5. Conclusions

We envision this relevance score as a starting point for future work in understanding the landscape of past and current oncological treatment. Essentially, the relevance score creates an *expectation* of certain treatment patterns, which can then be confirmed or refuted based on real-world data *observations*. The current model contains many scoring heuristics based on the experience of the authors. In future work we hope to explore machine learning approaches combined with additional annotation of a subset of our database to better refine the weight for each of the different factors associated with real world evidence of relevance. We also hope to deploy this model within the HemOnc.org framework, to allow visitors to not only have a quantitative relevance score associated with each treatment, but also the raw data from which this score is calculated.

References

[1] Arora S, Narayan P, Osgood CL, Wedam S, Prowell TM, Gao JJ, Shah M, Krol D, Wahby S, Royce M, Ghosh S, Philip R, Ison G, Berman T, Brus C, Bloomquist EW, Fiero MH, Tang S, Pazdur R, Ibrahim A, Amiri-Kordestani L, Beaver JA. U.S. FDA Drug Approvals for Breast Cancer: A Decade in Review. Clin Cancer Res. 2022 Mar 15;28(6):1072-86, doi: 10.1158/1078-0432.CCR-21-2600.

[2] Warner JL. Grappling with the Data Explosion in Oncology. Oncol Hematol Rev. 2015;11(102).

[3] Warner JL, Cowan AJ, Hall AC, Yang PC. HemOnc.org: A collaborative online knowledge platform for oncology professionals. J Oncol Pract. 2015 May;11(3):e336-50, doi: 10.1200/JOP.2014.001511.

[4] Esfahani K, Roudaia L, Buhlaiga N, Del Rincon SV, Papneja N, Miller WH Jr. A review of cancer immunotherapy: from the past, to the present, to the future. Curr Oncol. 2020 Apr;27(Suppl 2):S87-97, doi: 10.3747/co.27.5223.

[5] Farber S, Diamond LK, Mercer RD, Sylvester RF, Wolff JA. Temporary remissions in acute leukemia in children produced by folic acid antagonist, 4-aminopteroyl-glutamic acid. N Engl J Med. 1948 Jun;238(23):787-93, doi: 10.1056/NEJM194806032382301.

[6] Coiffier B, Lepage E, Brière J, Herbrecht R, Tilly H, Bouabdallah R, Morel P, Van Den Neste E, Salles G, Gaulard P, Reyes F. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. N Engl J Med. 2002 Jan;346(4):235-42, doi: 10.1056/NEJMoa011795.

[7] Butte AJ, Kohane IS. Unsupervised knowledge discovery in medical databases using relevance networks. Proc AMIA Symp. 1999:711-5.